

HOPE: A Python Just-In-Time compiler for astrophysical computations

Joël Akeret*, Lukas Gamper, Adam Amara, Alexandre Refregier

ETH Zurich, Institute for Astronomy, Department of Physics, Wolfgang Pauli Strasse 27, 8093 Zurich, Switzerland

Abstract

The Python programming language is becoming increasingly popular for scientific applications due to its simplicity, versatility, and the broad range of its libraries. A drawback of this dynamic language, however, is its low runtime performance which limits its applicability for large simulations and for the analysis of large data sets, as is common in astrophysics and cosmology. While various frameworks have been developed to address this limitation, most focus on covering the complete language set, and either force the user to alter the code or are not able to reach the full speed of an optimised native compiled language. In order to combine the ease of Python and the speed of C++, we developed HOPE, a specialised Python just-in-time (JIT) compiler designed for numerical astrophysical applications. HOPE focuses on a subset of the language and is able to translate Python code into C++ while performing numerical optimisation on mathematical expressions at runtime. To enable the JIT compilation, the user only needs to add a decorator to the function definition. We assess the performance of HOPE by performing a series of benchmarks and compare its execution speed with that of plain Python, C++ and the other existing frameworks. We find that HOPE improves the performance compared to plain Python by a factor of 2 to 120, achieves speeds comparable to that of C++, and often exceeds the speed of the existing solutions. We discuss the differences between HOPE and the other frameworks, as well as future extensions of its capabilities. The fully documented HOPE package is available at <http://hope.phys.ethz.ch> and is published under the GPLv3 license on PyPI and GitHub.

Keywords:

Python, Just-in-time compiler, benchmark

1. Introduction

In recent years, the Python programming language has gained a wide acceptance beyond its original use of simple scripting for system administration and test automatisation. Python has evolved to become a primary programming language for various industries and many scientific research fields, notably in astrophysics [1, 2]. The reasons for this success are the simplicity, flexibility and maintainability of Python code as well the wide range of its libraries. In particular, the widely used numerical Python packages NumPy[3] and SciPy[4] allow for fast prototyping and development of new applications. The flexible runtime and the dynamic typing are further features of the interpreter language that are valued by developers. However, these advantages come with a drawback: typically the execution time of Python programs can be slower than native compiled languages such as C or Fortran by orders of magnitudes.

While for many applications the performance of the software is not a priority, in astrophysics and cosmology where large simulations and data analysis over large data sets are often required, speed can be crucial (see e.g. [5, 6] and references therein for cosmology). In order, to increase the overall performance one can parallelize the programs to take advantage of multicore CPU architectures. An alternative and complementary approach is to focus on improving the single thread performance. The problem of improving the single thread performance of dynamic languages, such as Python, has been addressed in various different ways that can be grouped broadly into two categories: 1)

*Corresponding author

Email address: jakeret@phys.ethz.ch (Joël Akeret)

development of optimizing just-in-time (JIT) compilers and 2) development of faster interpreters [7]. The concept of the latter is to reduce the overhead introduced by the dynamic nature of the language. The idea of a JIT compiler is to produce faster machine or byte code during runtime when needed [8]. In the Python landscape both approaches have been implemented in various projects e.g. the PyPy¹ interpreter, the Numba² JIT package or the Cython³ C-extension and others.

While most of these approaches aim to support the complete language or a large portion of it, we find that some of the solutions are often intrusive (i.e. require the user to tailor the code to the framework) and that only a few of them are able to reach the full speed of an optimized native compiled C++ code. To fully combine the ease of Python and the speed of C++, we have therefore developed the HOPE package. HOPE is a specialized Python JIT compiler that supports a subset of the Python language - primarily numerical features commonly used in astrophysical calculations - and aims to reach the highest possible execution speed. The package translates Python code into C++ and is able to perform numerical optimization on mathematical expression at runtime. By using HOPE, the user benefits from being able to write common numerical code in Python while having the performance of compiled implementations. To enable the HOPE JIT compilation, the user only needs to add a decorator to the function definition. The package does not require additional information, which ensures that HOPE is as non-intrusive as possible.

We used the HOPE package in astrophysics applications such as the development of PyCosmo [9] and to rewrite UFig (the Ultra fast image generator) [10] into Python. PyCosmo is a Python cosmology package that numerically integrates the Einstein-Boltzmann differential equations and computes various cosmological observables. The HOPE package allowed us to improve the performance of the integration by a factor of $50 \times$ compared to the pure Python implementation by JIT compiling the integrated function. UFig is an image generator developed to simulate wide field cosmological galaxy surveys thus enabling tight control of systematics effects through forward modelling [5, 10]. HOPE has allowed us to rewrite the original C++ implementation of UFig into Python. Benchmarks show a performance comparable to the earlier C++ code, but the new implementation has increased the modularity, extensibility and readability of the code.

This paper is organised as follows. In section 2, we discuss the reason for the lower execution speed of dynamic languages, and review the existing solutions to address the performance implications in Python. Section 3 explains how we address the performance requirements using our JIT compiler package and describe the HOPE architecture and design. In section 4 we introduce the benchmarks we used to compare the performance of HOPE to the existing solutions. We show and discuss results in sections 5 and 6 and finally conclude in section 7. Information for downloading the HOPE package and on its performance on an alternative platform are described in Appendix A and B, respectively.

2. Review of existing solutions

Python is a dynamic interpreted language, which requires an interpreter for the execution of a program. Typically the performance of Python is much slower than C or comparable compiled languages. There are several reasons for this: Python has not been designed to be a fast language, but, instead, readability has been defined to be more important. In addition, everything in Python, including simple values, are represented as objects in memory. Therefore, the evaluation of a numerical expression requires the interpreter to unbox every value and to load the appropriate operation, as operators can be overwritten dynamically.

Currently four main implementation of interpreters exist: CPython, Jython, IronPython and PyPy. CPython is the reference implementation written in C and the most widely used Python interpreter. Jython and IronPython are alternative implementations of Python interpreters targeting the Java virtual machine and the .NET frameworks, but they typically do not improve performance[8]. PyPy is the most prominent effort to develop an alternative interpreter in order to increase the execution speed. In the following, we always refer to the CPython implementation if not explicitly stated otherwise.

If performance is critical, one approach is to write the critical code in C or C++ and interface the implementation with Python bindings. With this approach the performance requirements can typically be met but often at the cost of

¹<http://www.pypy.org>

²<http://numba.pydata.org>

³<http://www.cython.org>

Table 1: Description of existing packages with the version number used for the benchmarks.

Name	Description	Version
Numba	Numba is an open source package, which brings JIT compiling to Python by generating machine code using the LLVM software stack. Numba has been targeted to integrate into scientific Python applications. Similarly to HOPE the user can use a simple decorator to instruct Numba to compile a function. As JIT compiler Numba is able to use the type information available at runtime to generate the byte code.	0.13.3
Cython	Cython is a compiler for Python and for the Cython programming language. It allows the user to integrate C or C++ function into Python code as well as to compile Python code into machine code by adding static type declaration. This is done by using function decorators and type definition extensions. Cython supports the functionality of the NumPy package.	0.20.2
PyPy	PyPy is a Python interpreter written in a restricted subset of the Python language itself (RPython). PyPy is targeted towards speed and efficiency. The included JIT compiler allows translating the Python code into native code at runtime.	2.3.1
Nuitka	Nuitka is a static Python compiler aiming to support the entire Python language definition. It allows the compilation of a Python code into an executable.	0.5.3
numexpr	The numexpr package is designed for the efficient evaluation of numerical operations on array-like objects of the NumPy package. Additionally, the evaluation can be executed in parallel.	2.4

the readability. Experience shows that this makes it difficult for diverse users to maintain and extend the C/C++ code. A further drawback of this approach is that every change in the code requires a recompilation. During development this has to be performed repeatedly, which lowers the productivity and the development velocity of the team.

To be able to implement a code in Python but nevertheless be able to properly address the performance needs, different alternative approaches exist, including: Numba, Cython, Nuitka and numexpr (see Table 1). The Numba package is a JIT compiler that generates optimized machine code at runtime using the LLVM compiler infrastructure. A similar but different approach is to translate and compile the Python source code. The Cython C-extension for Python is widely used for this. The Nuitka project⁴ is a static Python compiler that focuses on full support of the language. A further approach is the numexpr package⁵ that specialises in evaluating numerical expressions and additionally allows parallel execution on multiple cores. Beside the frameworks listed in Table 1 further solutions exist. Pyston⁶ is a very new project under current development at Dropbox and is also built on the LLVM compiler infrastructure. Further frameworks are topic to research projects such as Pythran[11, 12], a static Python compiler and parakeet[13], which follows a similar approach as HOPE.

3. HOPE - A Python just-in-time compiler

HOPE is a specialized method-at-a-time JIT compiler written in Python. It translates Python source code into C++ and compiles the generated code at runtime. In contrast to other existing JIT compilers, which are designed for general purpose, we have focused our development of the subset of the Python language that is most relevant for

⁴<http://www.nuitka.net>

⁵<https://github.com/pydata/numexpr>

⁶<https://github.com/dropbox/pyston>

astrophysical calculations. By concentrating on this subset, HOPE is able to achieve very high performance for these applications.

HOPE is able to translate commonly used unary, binary and comparison operators as well as augmented assign statements. Currently the supported native built-in data types are `bool`, `int` and `float` and their corresponding NumPy data types (e.g. `int32`, `float32` resp. `int64`, `float64` etc). This applies to scalar values as well as the NumPy arrays with these types. Several features of NumPy and mathematical functions like *sin*, *cos*, *exp* etc. are also supported. HOPE, for instance, allows operations on NumPy arrays with the common slicing syntax (e.g. `a[5:, 5:] = x`). A full list of the functionality backed by the package is published online⁷ and a set of examples is provided on GitHub⁸. The package can be used with Python 2.7 as well as with Python 3.3+. We have tested HOPE in Linux and Mac OSX environment with `gcc` and `clang` compilers.

HOPE infers the data types by analysing the function signature and inspecting the Abstract Syntax Tree (AST) of the function at runtime. This information is used to translate the Python function into statically typed C++ code. Using Python's native extensions for C API⁹ the code is compiled into a shared object library, which is then loaded into the interpreter runtime. During the process of generating the C++ code, HOPE is able to apply various numerical optimisations in order to improve the execution performance of the compiled function (see section 3.1). Once compiled, the function is cached to disk to minimise lag on future executions of the same code.

We have chosen to generate C++ code and compiling the code into a library over other approaches, such as the direct generation of byte code. This design decision was made as the intermediate product - the generated code - is human readable, which greatly simplifies the development and debugging process and, furthermore, allows the use of automatic hardware specific optimisation of modern compilers without additional effort. The just-in-time compiling process is described in Fig. 1. A function call undergoes the following several steps:

Start The Python interpreter loads a function or method previously decorated with the `@hope.jit` decorator.

Cache verification HOPE checks if a compiled version of the requested functions has previously been cached. In case the code is executed the first time, HOPE returns a wrapper function containing a reference to the original function.

Parse function The first time the decorated function is called, the wrapper generates an abstract syntax tree (AST) by parsing the function definition using the Python built-in `ast` package.

Generate HOPE AST Using the visitor pattern, the Python AST is traversed and a corresponding HOPE specific AST is generated. During the traversal we use the Python built-in `inspect` package to infer the data types of the live objects such as parameters, variable and return values. Using this information we are able to statically type the HOPE AST i.e. scalar variables will be assigned a type and array like variable will receive a data type and a shape information. Operations on arrays with the same shape will be grouped into blocks. Furthermore a scope is assigned to each variable in order to identify if it will be passed as parameter or if it has to be instantiated. In the latter case HOPE distinguishes between temporary variables, used only once (block scope), and variables used multiple time in the function (body scope).

Numerical optimization HOPE traverses the new AST in order to identify numerical optimisation possibilities and alters the tree accordingly. A detailed explanation of the optimization is discussed in section 3.1.

Generate C++ code A C++ code is generated from the HOPE AST. First the function signature is created using the name of the function, the type of the return value and the names of the parameters including their type and shape information. If necessary, statements for the instantiation of new variables are generated. Next, each block is turned into a loop statement according to the shape information of the contained arrays. By grouping array operations, we are able to evaluate the operation element-wise, which improves cache locality. For variables with block scope we can avoid allocating a whole array and instead a scalar value can be allocated (an example is discussed in section 3.2). Finally, the generated code is augmented with Python's native extensions API statements, so that the code can be called from the Python interpreter.

⁷<http://pythonhosted.org/hope/lang.html>

⁸<https://github.com/cosmo-ethz/hope/tree/master/examples>

⁹<https://docs.python.org/2/extending/extending.html>

Compile code to shared object library The Python built-in `setuptools` package is then used to compile a shared object library from the generated code. Using this package and defining the generated code as an extension greatly simplifies the handling of compiler and ensures compatibility with the interpreter.

Add library to cache Using the extracted information from the function signature and a hash over the function body the compiled shared object library is cached for future calls.

Load library The shared object library is dynamically loaded into the runtime and the pointer to the wrapper is replaced with the pointer to the function in the shared object library to avoid unnecessary overhead.

Execute compiled function A call to the function is directed to the function in the shared object library and executed with the passed parameters.

Subsequent function call HOPE analyses the types of the passed arguments and queries the cache for a function matching the requested name and arguments. If the system registers a cache hit for a function, the shared object library is then loaded into the runtime and the compiled function is evaluated, otherwise a new function will be generated and compiled.

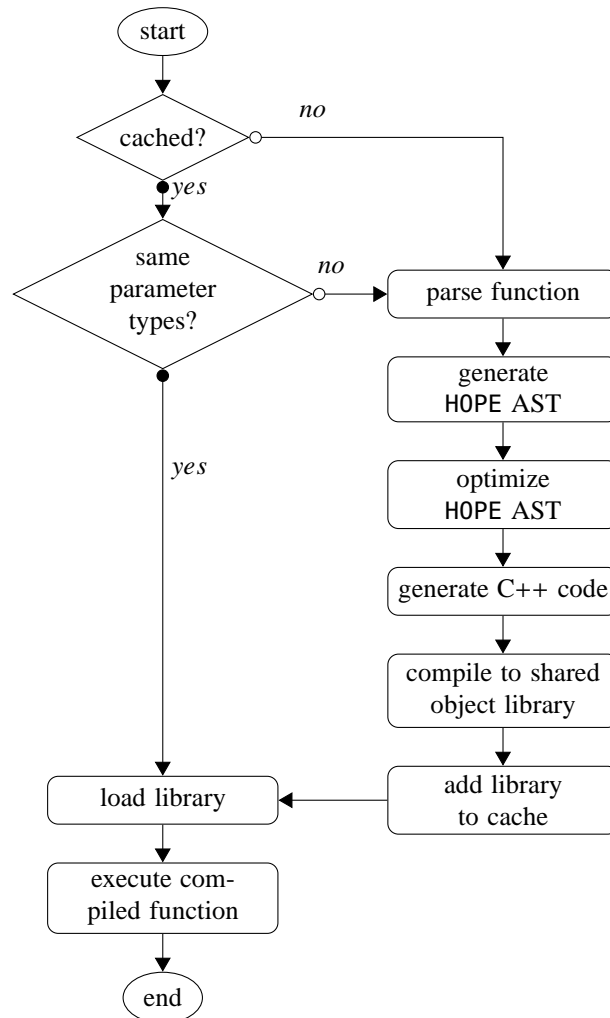


Figure 1: Flow diagram of the step executed during a function call decorated with HOPE

3.1. Optimisation

After the HOPE specific AST has been created the package performs a static recursive analysis of the expressions to introduce numerical optimisation. The supported possibilities are divided into three groups: 1) simplification of expressions, 2) factorising out subexpressions and 3) replacing the pow function for integer exponents. To simplify expression we have used the SymPy library[14]. SymPy is a Python library for symbolic mathematics and has been entirely written in Python. To apply the optimization, the AST expression is translated into SymPy syntax AST and passed to the `simplify` function. The function applies various different heuristics to reduce the complexity of the passed expression. The simplification is not exactly defined and varies depending on the input. For instance, one example of simplification is that $\sin(x)^2 + \cos(x)^2$ will be simplified to 1. Furthermore the SymPy library is used to factorise out recurring subexpression (common subexpression elimination) using the previously created SymPy AST and SymPy's `cse` function.

From C++11 on, the pow function in the C standard library is not overloaded for integer exponents¹⁰. The internal implementation of the computation of a base to the power of a double exponent is typically done using a series expansion, though this may vary depending on the compiler and hardware architecture. Generally this is efficient for double exponents but not necessarily for integer exponents. HOPE therefore tries to identify power expressions with integer exponents and factorizes the expression into several multiplications e.g. $y = x^5$ will be decomposed into $x_2 = x^2$ and $y = x_2 \times x_2 \times x$. This reduces the computational costs and increases the performance of the execution.

3.2. Example

In this section, we show the translation process of HOPE using the following simple example:

```
@hope.jit
def fkt(x,y):
    z = x**2 + 1
    return z + y
```

As soon as this function is called, HOPE will create a statically typed AST using the type information available at runtime. Assuming that the function call was done with `x` and `y` defined as one dimensional `float64` NumPy array with same length, the resulting AST can be visualised as the following pseudo code:

```
fkt(numpy::float64[] x, numpy::float64[] y) -> numpy::float64[] {
    new numpy::float64[] ret = numpy::float64[ len(x) ]
    for(int i = 0; i < len(x); ++i){
        numpy::float64 tmp = x[i] * x[i]
        numpy::float64 z = tmp + 1
        ret[i] = z + y[i]
    }
    return ret
}
```

Using the inspected data types of the parameters and by analysing the mathematical expressions, HOPE is able to identify that operations are performed on arrays with compatible data types and dimensions and will group those expressions into a block. The result of this block is a new array with the same data type and dimension, which will also be used as return value. This information tells HOPE that a new array has to be instantiated and how the function signature will be defined. Next, the block is turned into a loop over every element of the arrays, which includes the power operation as well as the addition operations. Finally, the power operation on the `x` value with the integer exponent is optimised into a multiplication of the `x` value.

3.3. Quality assurance

The HOPE package has been developed using the test-driven development (TDD) approach, allowing us to ensure a high level of code quality and numerical accuracy up to type-precision. For every supported language feature we have written a unit test that is executed for all the supported data types and array shape combinations. This results in over 1600 unit tests being executed on our continuous integration server per supported Python version.

¹⁰<http://en.cppreference.com/w/cpp/numeric/math/pow>

4. Benchmarks

In this section, we describe benchmarks aimed at assessing the performance of HOPE and at comparing it to the other packages described in section 2. Since a standardised set of benchmarks for testing the performance of software does not exist, we have generated our own series of tests. The first set are three numerical micro-benchmarks published on the Julia language website¹¹ and which are supported by the current version of HOPE. These benchmarks have already been applied to various languages. We omitted benchmarks with matrix manipulations, as good performance typically depend on the proper use of specialised matrix libraries such as Intel's Math Kernel Library (MKL) or the Basic Linear Algebra Subprograms (BLAS). Additionally, we defined two benchmarks favoring numexpr as they test the ability to compute polynomials and mathematical expression which can be simplified. Finally, we have specified two special purpose benchmarks representing simplified versions of common problems in astrophysics to compare the performance of the general-purpose frameworks. To have a reference and baseline, we additionally implemented all the benchmarks in C++.

For our benchmark tests we have chosen packages where the user should, in principle, be able to write a code in Python and that only minor modifications would be needed to enable the performance improvements. All the benchmarks have been made available online as IPython notebooks¹². The following describes the benchmarks and provide the associated Python source code.

4.1. Fibonacci sequence

The Fibonacci sequence is a common measurement of the execution speed of repeated recursive function calls. The sequence is defined as:

$$F_n = F_{n-1} + F_{n-2}, F_0 = 0, F_1 = 1.$$

We executed all the benchmarks with $n = 20$ resulting in the answer 6765. The Python implementation is as follows:

```
def fib(n):
    if n < 2:
        return n
    return fib(n - 1) + fib(n - 2)
```

4.2. Quicksort

Quicksort is a powerful yet simple sorting algorithm that gained widespread adoption in Unix as the default sorting function and in the C standard library. It extends the complexity of the previous recursive benchmark with a loop, multiple comparison and different data types. The algorithm was implemented as follows:

```
def qsort_kernel(a, lo, hi):
    i = lo
    j = hi
    if False: return a
    while i < hi:
        pivot = a[(lo+hi) // 2]
        while i <= j:
            while a[i] < pivot:
                i += 1
            while a[j] > pivot:
                j -= 1
            if i <= j:
                tmp = a[i]
                a[i] = a[j]
                a[j] = tmp
                i += 1
                j -= 1
        if lo < j:
            qsort_kernel(a, lo, j)
    lo = i
    j = hi
    return a
```

¹¹<http://julialang.org>

¹²<http://hope.phys.ethz.ch>

The algorithm is used in the benchmarks to sort an array of 5'000 uniform random `float64`.

4.3. *Pi sum*

This benchmark is a simple approximation of $\pi/2$, testing the run time behavior of two nested for-loops.

$$p = \sum_{j=1}^{501} \sum_{k=1}^{10001} \frac{1}{k * k}$$

```
def pi_sum():
    for j in range(1, 501):
        sum = 0.0
        for k in range(1, 10001):
            sum += 1.0/(k*k)
    return sum
```

4.4. *10th order polynomial*

In this numerical benchmark a logarithm is approximated using a Taylor expansion resulting in a 10th order polynomial. The approximation is defined as follows:

$$\ln(x) \approx \sum_{i=1}^9 = (-1)^{i-1} \frac{(x-1)^i}{i}$$

and can be implemented in Python as such:

```
def ln_python(X):
    return (X-1) - (X-1)**2 / 2 + (X-1)**3 / 3 - (X-1)**4 / 4 +
            (X-1)**5 / 5 - (X-1)**6 / 6 + (X-1)**7 / 7 -
            (X-1)**8 / 8 + (X-1)**9 / 9
```

For our benchmarks, a slightly optimized version of this implementation has been used:

```
def ln_python_exp(X):
    x = (X - 1)
    x2 = x*x
    x4 = x2*x2
    x6 = x4*x2
    x8 = x4*x4
    return x - x2 / 2 + x * x2 / 3 - x4 / 4 + x * x4 / 5 - x6 / 6 + x6 * x / 7 - x8 / 8 + x8 * x / 9
```

where X as been defined as an array of 10'000 uniform random `float64`.

4.5. *Simplify*

This benchmark has been specified as

$$y(x) = \sin^2(x) + \frac{(x^3 + x^2 - x - 1)}{(x^2 + 2 * x + 1)} + \cos^2(x)$$

```
def y(X):
    return np.sin(x)**2 + (x**3 + x**2 - x - 1)/(x**2 + 2*x + 1) + np.cos(x)**2
```

Where X as been defined as an array of 5'000 uniform random `float64`. The benchmark tests the ability of the packages to efficiently compute polynomial expressions. As the expression can be simplified to $y = x$ this benchmark will favour frameworks, which analyse and optimise the expression such as `numexpr` and `HOPE`.

4.6. Pairwise distance

Computing the two-point correlation function for a set of points is a common problem in many areas of astronomy and cosmology. A similar but simplified problem is the computation of the distance between all the points. Using the l^2 norm this results in a $n \times n$ matrix for a given input array of size n . The distance is defined as:

$$D_{i,j} = \sqrt{\sum_{k=0}^N (X_{i,k} - X_{j,k})^2}, \forall i \in I, \forall j \in J$$

where $N = 3$ is the number of dimensions and $I = J = 1000$ is the number of points in the array. The naive solution used in the benchmark is implemented as follows:

```
def pairwise_python(X, D):
    M = X.shape[0]
    N = X.shape[1]
    for i in range(M):
        for j in range(M):
            d = 0.0
            for k in range(N):
                tmp = X[i, k] - X[j, k]
                d += tmp * tmp
            D[i, j] = np.sqrt(d)
```

Additionally we benchmark the vectorized implementation of this test using the NumPy package:

```
def pairwise_numpy(X, D):
    M = X.shape[0]
    for i in range(M):
        D[i, :] = np.sqrt(np.sum((X[i, :] - X[:, :]) ** 2, axis=1))
```

4.7. Point Spread Function

This benchmark has been inspired by calculations that simulate astronomical imaging data. One step in these simulations is to make realistic simulations of the images of stars. For many astronomical applications stars are smaller than the resolution of the instrument. Therefore, the resulting images are realisations of the Point Spread Function (PSF) coming from the finite resolution and atmospheric effects. A good model for the PSF of ground-based telescopes is a circular Moffat profile [15], given by

$$I(r) = I_0 \left[1 + \left(\frac{r}{\alpha} \right)^2 \right]^{-\beta}$$

where I_0 is the value at the origin ($r=0$), α and β are parameters depending on the conditions of the observation (see Appendix B in [10]). The numerical integration in the x and y direction is done using Gauss-Legendre integration with 7th order coefficients:

```
def pdf(density, x_range, y_range, x_center, y_center, w2D, r50, b, a):
    for x in range(x_range):
        for y in range(y_range):
            dr = np.sqrt((x - x_center) ** 2 + (y - y_center) ** 2)
            density[x, y] = np.sum(w2D * 2 * (b - 1) / (2 * np.pi * (r50 * a)**2) * (1 + (dr / (r50 * a))**2)**(-b))
    return density
```

The density and w2D variables are 20×20 and 7×7 Numpy arrays, respectively, while the other parameter are defined as scalar values. A very similar approach has been implemented in the UFig project [10].

5. Results

We ran the benchmarks on a MacBook Pro OS X 10.9.2 with a Intel Core i7 2.3 Ghz processor and 16 GB 1600 MHz DDR3 memory as well as on a Mac Pro (see Appendix B for details). The results with the second system can be found in the Appendix B. Table 1 shows the versions of the packages used to conduct the test runs. To compile

the C++ code and HOPE benchmarks we used `clang-503.0.40` and the `-march=native`, `-stdlib=libc++` and `-std=c++11` compiler options.

The benchmarks from section 4 have been executed with all the frameworks of Table 1 except `numexpr` which could only be used for the computation of the 10^{th} order polynomial and the `simplify` benchmark. Every test has been executed 100 times using the built-in `timeit` package and the median of the measured runtime has been used for comparison in order to reduce the influence of other processes interfering with the timing accuracy.

For all the benchmarks using Cython we have disabled the `wraparound` and `boundscheck` compiler flags. The benchmarks with `numexpr` have been executed with parallelization using the 8 cores available on the testing infrastructure. Table 2 shows the benchmark run times relative to the C++ implementation. The Pairwise distance benchmark has been performed using both, the plain Python implementation, as well as the vectorized implementation using the NumPy package (see results in parenthesis in the table).

Generally Cython was able to improve the performance by a factor $1.2\times$ up to $53\times$ compared to the Python runs. An exception is the `simplify` benchmark where the timing was marginally worse. The runs with the Numba package show that, if the framework is able to properly infer the data types and compile the function to LLVM, the performance is comparable to Cython. We have not been able to perform the Quicksort benchmark using Numba as the execution resulted in an `Internal error`. This might be due to the currently limited support for recursive function calls. The Pairwise distance benchmark shows that vectorizing the code with the NumPy package can drastically improve performance. The impact is more pronounced on Python and Nuitka ($155\times$ and $131\times$, respectively) as with PyPy and NumPy ($4.6\times$). However, vectorisation is not always possible or can be difficult and may increase memory consumption due to vector broadcasting. The alternative interpreter PyPy was able to improve the performance compared to CPython in the first three benchmarks ($1.7\times$ - $26\times$) but performed worse in the others ($1.2\times$ - $17\times$ slower). `numexpr` achieved better runtimes than Python for the two benchmarks conducted. The improvement of the parallelisation was small, which may be ascribable to the parallelisation overhead. For larger problems the speed up could be larger. Attempts to conduct the benchmarks with the `parakeet` project, which is implementing a similar approach to HOPE, succeeded only for the `Simplify` and `Pairwise distance` benchmark. The improvement for the first benchmark was small and for the second, `parakeet` was faster than the vectorized NumPy implementation by a factor of $5.1\times$.

Our HOPE package was able to speed up the computation in all of the benchmarks. A very large improvement can be seen in the pairwise distance benchmark. This can be ascribed to the naive Python implementation using multiple nested loops. As the bad performance of loops in Python is commonly known, we expect developers to implement this benchmark using Numpy instead. Therefore we disregard this speed up for the further comparisons. As can be seen in Table 2 HOPE improved the performance by a factor of $2.4\times$ to $119\times$ compared to the Python runs and was only marginally slower than the native C++ implementation.

For the Pi sum benchmark Numba and PyPy were able to improve the performance as much as HOPE and to match that of C++. In the Pairwise distance benchmark Cython outperformed HOPE and reached the same performance as the native C++ implementation. The `simplify` benchmark shows the power of the optimisation capabilities of the HOPE package. Disabling the optimisation option in HOPE would result in similar timing as the native C++ implementation. For specialised problems such as the Point Spread Function, HOPE and C++ both clearly outperform all the alternative frameworks tested. The benchmarks conducted on the Mac Pro system yield comparable results (see Table B.3 of Appendix B).

6. Discussion

When using Cython, the user has to provide to the package a statically typed function signature, as well as statically typed variables in order to achieve the desired performance improvements. The package `numexpr` yields good performance (as can be seen in Table 2) but it requires the user to write expressions as strings. This limits the applicability and, more importantly, removes syntax highlighting and variable recognition in editors and integrated development environments (IDE). The benchmarks indicate that the PyPy interpreter is only able to partially increase the performance compared to CPython. The project is under active development and interesting concepts are being addressed within it. Packages with C extension such as SciPy are currently not fully supported, which limits the use of PyPy in scientific applications. Since compiling the Python code with the static compiler Nuitka improved the performance only marginally, the overhead arising from the compilation process do not appear to be justified.

Table 2: Benchmarks times relative to C++. Best results are highlighted in bold.

	Python (NumPy)	Numba	Cython	Nuitka (NumPy)	PyPy (NumPy)	numexpr (8 cores)	HOPE	C++
Fibonacci	57.4	65.7 ^a	1.1	26.7	21.1	—	1.1	1.0
Quicksort	79.4	— ^b	4.6	61.0	45.8	—	1.1	1.0
Pi sum	27.2	1.0	1.1	13.0	1.0	—	1.0	1.0
10 th order	2.6	2.2	2.1	1.2	12.1	1.4	1.1	1.0
Simplify	1.4	1.5 ^{ab}	1.8	1.4	23.2	0.6	0.015	1.0
Pairwise distance	1357.8 (8.7)	1.8	1.0	1247.7 (9.5)	277.8 (60.4)	—	1.7	1.0
Star PSF	265.4	250.4 ^a	46.2	234.6	339.5	—	2.2	1.0

^a Numba was not able to compile down to LLVM^b Compilation attempt resulted in Internal error

Numba, which is the only package besides HOPE that does not require the user to alter the code or change the runtime environment, shows good performance as soon the package is able to compile down to LLVM. As the project is also under active development we expect that further support and features will be implemented soon. Focusing on a subset of the Python language enables HOPE to generate C++ code targeted towards high execution performance without the need for the user to modify the Python implementation. The performance differences compared to the C++ implementation arise through small overheads introduced by the code generation process. It has to be noted that HOPE is still under active development and many language features of Python are currently not supported. In cases where HOPE is not able to translate the code it will provide the user the according information including the line of code, which caused the problem. In these cases exploring possible improvements in performance through Cython would be an option. This requires the user manually adapt the code. For an experienced user gains are likely to be possible.

7. Conclusion

Python is becoming increasingly popular in the science community due to the large variety of freely available packages and the simplicity and versatility of the language. However, a drawback of Python is the low runtime and execution performance of the language. For many use cases this is acceptable but for large simulations and numerical computations, such as those used in astrophysics and cosmology, accelerating the performance of codes is crucial. This can be achieved by parallelizing the computation on multicore CPU architectures. Alternatively and complementarily, the single thread performance of the code can be optimized. Rewriting the application in C (or other compiled languages) can be time consuming and reduces the readability and maintainability of the code. A set of solutions exists in the Python landscape to improve the performance, such as alternative interpreters, static Python code compiler or just-in-time compilers. We find that some those solution are intrusive i.e. they require the user to change the code and some are not able to fully achieve the speed of a corresponding C/C++ implementation.

To address these limitations, we introduced HOPE, a specialised Python just-in-time compiler able to apply numerical optimisation to mathematical expressions during the compilation process. We conducted different benchmarks to assess its performance and compared it with existing solutions. The tests show that HOPE is able to improve the performance compared to plain Python by a factor of $2.4\times$ - $119\times$ depending on the benchmark scenario. We find that the performance of our package is comparable to that of C++. Some of the other packages that we tested are also able to improve the execution speed but do not increase the performance in specialized test cases such as the computation of a ground-based point spread function. We have used our package to improve the performance of the PyCosmo project [9] as well as to be able to rewrite the Ultra fast image generator (UFig) C++ package[10] in Python without compromising its performance. We plan to apply HOPE to further projects and therefore continuously increase its supported language features and improve its optimization capabilities. To simplify the installation we are

distributing the code through the central PyPI server¹³ and provide the full documentation of the package online¹⁴. In Appendix A we describe the distribution and installation of HOPE.

¹³<https://pypi.python.org/pypi/hope>

¹⁴<http://hope.phys.ethz.ch>

Appendix A. Distribution

Detailed documentation, supported language subset and installation instructions can be found on the package website <http://hope.phys.ethz.ch>. The HOPE package is released under the GPLv3 license and has been uploaded to PyPI¹⁵ and can be installed using pip¹⁶:

```
$ pip install hope --user
```

This will install the package and all of the required dependencies. The development is coordinated on GitHub <http://github.com/cosmo-ethz/hope> and contributions are welcome.

Appendix B. Mac Pro results

The benchmarks described in section 4 have been conducted on a Mac Pro system with an Intel Xeon E5 Ivy Bridge 12-core (2.7 GHz) and 64GB (4x16GB) of 1866MHz DDR3 memory to test the effect of different hardware. As can be seen in Table B.3, the timings are comparable to the results in Table 2, leading to the same conclusions as discussed in section 6.

Table B.3: Benchmarks times relative to C++ (Mac Pro). Best results are highlighted in bold.

	Python (NumPy)	Numba	Cython	Nuitka (NumPy)	PyPy (NumPy)	numexpr (8 cores)	HOPE	C++
Fibonacci	65.5	62.6 ^a	1.0	31.6	12.4	—	1.0	1.0
Quicksort	109.0	— ^b	5.0	62.7	39.2	—	1.1	1.0
Pi sum	31.9	1.0	1.0	16.6	1.1	—	1.0	1.0
10 th order	4.0	3.9	3.8	3.5	25.8	2.3	1.2	1.0
Simplify	1.4	1.3 ^{ab}	1.4	1.2	7.9	0.8	0.017	1.0
Pairwise distance	2076.6 (13.9)	16.2	1.0	1460.8 (13.7)	254.4 (75.2)	—	1.7	1.0
Star PSF	374.7	385.9 ^a	98.8	348.7	500.1	—	2.0	1.0

^a Numba was not able to compile down to LLVM

^b Compilation attempt resulted in Internal error

¹⁵<https://pypi.python.org/pypi/hope>

¹⁶www.pip-installer.org/

References

- [1] T. Software, Tiobe index for august 2014 (08 2014) [cited 2014-08].
URL <http://www.tiobe.com/index.php/content/paperinfo/tpci/index.html>
- [2] D. N. R. J. Cass, S., Ieee spectrum's 2014 ranking (07 2014) [cited 2014-08].
URL <http://spectrum.ieee.org/static/interactive-the-top-programming-languages>
- [3] S. van der Walt, C. C., G. Varoquaux, The numpy array: A structure for efficient numerical computation, *Comput. Sci. Eng.* 13.
URL <http://dx.doi.org/10.1109/MCSE.2011.37>
- [4] E. Jones, T. Oliphant, P. Peterson, et al., SciPy: Open source scientific tools for Python, [Online; accessed 2014-08-12] (2001–).
URL <http://www.scipy.org/>
- [5] A. Refregier, A. Amara, A way forward for Cosmic Shear: Monte-Carlo Control Loops, *Physics of the Dark Universe* 3 (2014) 1–3. arXiv:1303.4739, doi:10.1016/j.dark.2014.01.002.
- [6] J. Akeret, S. Seehars, A. Amara, A. Refregier, A. Csillaghy, CosmoHammer: Cosmological parameter estimation with the MCMC Hammer, *Astronomy and Computing* 2 (0) (2013) 27 – 39. arXiv:1212.1721.
URL <http://www.sciencedirect.com/science/article/pii/S221313371300022X>
- [7] M. Arnold, S. Fink, D. Grove, M. Hind, P. Sweeney, A survey of adaptive optimization in virtual machines, *Proceedings of the IEEE* 93 (2) (2005) 449–466. doi:10.1109/JPROC.2004.840305.
- [8] A. Cuni, High performance implementation of python for cli/.net with jit compiler generation for dynamic languages, Ph.d. thesis, DISI, Università di Genova, technical Report DISI-TH-2010-05 (2010).
- [9] A. Refregier, A. Amara, L. Gamper, J. Akeret, PyCosmo: A Cosmological Boltzmann Solver in Accelerated Python, *Astronomy and Computing* (in. prep) tbd.
- [10] J. Bergé, L. Gamper, A. Réfrégier, A. Amara, An Ultra Fast Image Generator (UFIG) for wide-field astronomy, *Astronomy and Computing* 1 (2013) 23–32. arXiv:1209.1200, doi:10.1016/j.ascom.2013.01.001.
- [11] Pythran: Enabling Static Optimization of Scientific Pythran: Enabling Static Optimization of Scientific Python Programs, *PROC. OF THE 12th PYTHON IN SCIENCE CONF. (SCIPY 2013)*.
- [12] S. Guelton, J. Falcou, P. Brunet, Exploring the vectorization of python constructs using pythran and boost simd, in: *Proceedings of the 2014 Workshop on Programming Models for SIMD/Vector Processing, WPMVP '14*, ACM, New York, NY, USA, 2014, pp. 79–86. doi:10.1145/2568058.2568060.
URL <http://doi.acm.org/10.1145/2568058.2568060>
- [13] A. Rubinsteyn, E. Hielscher, W. N., S. D., Parakeet: A just-in-time parallel accelerator for python, in: *Presented as part of the 4th USENIX Workshop on Hot Topics in Parallelism, USENIX, Berkeley, CA, 2012*.
URL <https://www.usenix.org/conference/hotpar12/parakeet-just-time-parallel-accelerator-python>
- [14] SymPy Development Team, SymPy: Python library for symbolic mathematics (2014).
URL <http://www.sympy.org>
- [15] A. F. J. Moffat, A theoretical investigation of focal stellar images in the photographic emulsion and application to photographic photometry, *Astronomy and Astrophysics* 3 (1969) 455.